Talanta 139 (2015) 50–55

Contents lists available at ScienceDirect

Talanta

journal homepage: www.elsevier.com/locate/talanta

Digital image-based classification of biodiesel

Gean Bezerra Costa [a], David Douglas Sousa Fernandes [b], Valber Elias Almeida [b], Thomas Souto Policarpo Araújo [b], Jessica Priscila Melo [b], Paulo Henrique Gonçalves Dias Diniz [a,b], Germano Véras [a,b,*]

[a] Programa de Pós-Graduação em Ciências Agrárias, Universidade Estadual da Paraíba, 58.429-500 Campina Grande, PB, Brazil
[b] Departamento de Química, Centro de Ciências e Tecnologia, Universidade Estadual da Paraíba, 58.429-500 Campina Grande, PB, Brazil

ARTICLE INFO

Article history:
Received 3 December 2014
Received in revised form
22 February 2015
Accepted 23 February 2015
Available online 3 March 2015

Keywords:
Biofuel
Webcam
Color histograms
Pattern recognition
Successive Projections Algorithm

ABSTRACT

This work proposes a simple, rapid, inexpensive, and non-destructive methodology based on digital images and pattern recognition techniques for classification of biodiesel according to oil type (cottonseed, sunflower, corn, or soybean). For this, differing color histograms in RGB (extracted from digital images), HSI, Grayscale channels, and their combinations were used as analytical information, which was then statistically evaluated using Soft Independent Modeling by Class Analogy (SIMCA), Partial Least Squares Discriminant Analysis (PLS-DA), and variable selection using the Successive Projections Algorithm associated with Linear Discriminant Analysis (SPA-LDA). Despite good performances by the SIMCA and PLS-DA classification models, SPA-LDA provided better results (up to 95% for all approaches) in terms of accuracy, sensitivity, and specificity for both the training and test sets. The variables selected Successive Projections Algorithm clearly contained the information necessary for biodiesel type classification. This is important since a product may exhibit different properties, depending on the feedstock used. Such variations directly influence the quality, and consequently the price. Moreover, intrinsic advantages such as quick analysis, requiring no reagents, and a noteworthy reduction (the avoidance of chemical characterization) of waste generation, all contribute towards the primary objective of green chemistry.

1. Introduction

In worldwide fuel markets, there is an increasing interest in renewable fuels. In contrast to traditional petroleum-based diesel fuels that cause environmental degradation along the time, biofuels are a good alternative. Biofuels diversify sources of energy, and for being renewable, they are notably eco-friendly. They are free of heavy metals, and can either partially or totally substitute fossil fuels. Among biofuels, such as bioethanol, biogas, and bio-oil, biodiesel stands out as the most promising [1–5].

Since it is considered as an alternative to petroleum diesel in many countries, biodiesel is normally introduced as part of a pre-existing national energy matrix, for each country; according to their respective national energy, agricultural, and environmental policies, the diesels are blended in differing proportions. For instance, in Argentina diesel is blended to 5% or 20% biodiesel, while in Germany the consumer chooses the final biodiesel/diesel mixture ratio to be used. In Brazil, the National Agency of Petroleum, Natural Gas, and Biofuels (ANP) establishes regulations for quality control of biodiesel, and nationwide standards for the addition of biodiesel to diesel at 5% (B5) [6].

Amidst the verified growing demand for biodiesel, Brazilian and European policies require comprehensive searches of the various feedstocks, including vegetable oils and animal fats [7,8]. In Brazil, the main sources of biomass for biodiesel production are soybean, beef tallow, and cottonseed; other potential oil sources are also investigated [9].

An immediate consequence of higher biodiesel production and diversification of its feedstock is the new relevance which quality control assumes. The quality of biodiesels and/or biodiesel–diesel blends has been assessed for glycerol quantification, mono-, di-, triglycerides, methanol, water, Na, K, P, and steroids (in biodiesel), and using the trans-esterification reaction. Many analytical methodologies have been reviewed, including chromatographic, spectroscopic, physical properties-based, and wet chemical methods. However, these techniques are laborious, expensive, and destructive; they also require sample preparation using various chemicals, which generates waste [5,10,11].

Recently, studies have been published showing the importance of analytical methodologies to determine the starting oil for biodiesel production [1,12]. Near infrared (NIR) spectroscopy [1], and

* Corresponding author at: Departamento de Química, Centro de Ciências e Tecnologia, Universidade Estadual da Paraíba, 58.429-500 Campina Grande, PB, Brazil. Tel.: +55 83 3315 3300x3516.
E-mail address: germano.veras@pq.cnpq.br (G. Véras).

http://dx.doi.org/10.1016/j.talanta.2015.02.043
0039-9140/

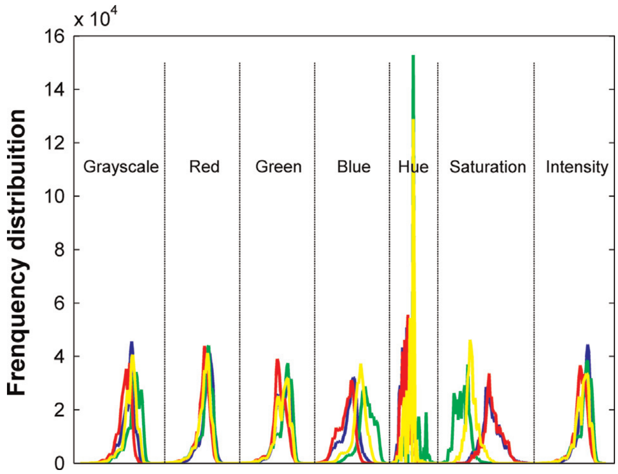

Fig. 1. Mean histograms in the Grayscale, Red, Green, Blue, Hue, Saturation, and Intensity channels for each class of biodiesel sample: cottonseed (blue line), corn (red line), sunflower (green line) and soybean (yellow line). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

ultraviolet–visible (UV–vis) spectroscopy [12], associated with Soft Independent Modeling of Class Analogy (SIMCA) have been used to classify biodiesel samples according to their oil sources. Screening of biodiesel/diesel mixtures with respect to their starting oils has also been investigated [13].

In this sense, it would be interesting to develop a simple methodology based on digital images and pattern recognition techniques for classification of biodiesel with respect to the oil source that is rapid, inexpensive, and non-destructive. Digital imaging has become increasingly more important due to its ability to perform fast and low-cost analyses [14–16]. Digital images can be obtained from many devices, including digital cameras, webcams, scanners, and even mobile phones [17–22]. The use of digital images can replace the human visual system, eliminating the subjective nature of the analysis, which is substantially influenced by environmental conditions and subject to inconsistencies. In order to promote a standardized specification, color systems define a three-dimensional coordinate space, where each color is represented by a single point [23,24]. Red–Green–Blue (RGB), Hue–Saturation–Intensity (HSI) and Grayscale have been the most widely used color systems [14–29].

For analytical purposes, digital images are decomposed into color histograms, which describe the statistical distribution of the pixels as a function of the recorded color component; they have been successfully used as input data for classification of teas [26], honeys [27], coffees [28], and bacteria [29]. SIMCA, Principal Component Analysis–Linear Discriminant Analysis (PCA–LDA), Partial Least Squares Discriminant Analysis (PLS-DA) and/or variable selection using the Successive Projections Algorithm associated with Linear Discriminant Analysis (SPA-LDA) have all been employed as multivariate classifiers.

In this work a new methodology is proposed based on digital images and pattern recognition techniques for classification of biodiesel according to oil type (cottonseed, sunflower, corn, and soybean). For this, differing color histograms in the RGB, HSI, Grayscale channels, and their combinations were extracted from the digital images and used as analytical information, and then statistically evaluated using SIMCA, PLS-DA, and SPA-LDA.

2. Material and methods

2.1. Samples

The present study involved a total of 120 samples (brands and lots) of different refined oils, comprising 30 samples each of: cottonseed, sunflower, corn, and soybean biodiesel. Methanol was used to promote the trans-esterification reaction catalyzed by potassium hydroxide. The biodiesel samples obtained were continuously washed with distilled water and hydrochloric acid at 0.1 mol/L until reaching a pH of 7.0. The washing step helps to remove visible residues, such as glycerin, methanol, fatty acids, salts, and catalyst excesses [1].

For image acquisition (see Section 2.3), 5.0 mL of each biodiesel sample was used to fill a Petri plate in order to promote uniformity, and to maintain the same overall visual characteristics of the sample surface.

2.2. Apparatus

The apparatus built for image acquisition was described by Almeida and coworkers [29]. A 30 cm × 22 cm × 23 cm wood compartment was built in order to isolate the biodiesel samples from external light, ensuring reproducibility of the images' capture process, i.e. both quality and uniformity of the recorded images. A Petri plate (used as a sample holder), and a Webcam, Microsoft® model Lifecam Cinema, with 7.1 megapixels, were strategically disposed inside the compartment for image acquisition. In addition, the compartment was internally coated with white paper. An 8 W commercial fluorescent lamp was placed outside the compartment at 25 cm above the sample, in order to provide uniform internal illumination, in accordance with optical principles.

2.3. Image acquisition

Image acquisition takes into account the overall visual characteristics of the sample surface. The images acquired were stored in JPEG format, which is a compressed format. However, this does not affect the classification results because the images acquired are composed of 16.7 million colors with a resolution of 2880 × 1620 pixels. Since color histograms describe the statistical distribution of the pixels as a function of the recorded color component, the selection of the region of interest (ROI) provides enough information to obtain a suitable classification, as discussed in Section 3.2. This is corroborated by the good results obtained in other analytical applications, as described elsewhere [26–29].

Five sequential images were captured for each sample, and then decomposed into color histograms from the ROI selection. Mean histograms for each sample were calculated and used as the instrumental response. The data matrix was then formed by the samples located in rows, while the columns correspond to the color levels.

2.4. Color histograms and data analysis

Color histograms describing the statistical distribution of the pixels as a function of the recorded color component in the RGB, HSI and Grayscale channels were obtained from each digital image using a routine written in Matlab R2011b. For this, an ROI of 1008 × 567 pixels was selected from the center of each digital image, containing approximately 35% of the total image area. Using the ROI alone, histograms employing Red, Green, Blue, Hue, Saturation, Intensity, and Grayscale were built and used as analytical information. RGB components, and Grayscale levels vary from 0 to 255 (256 levels); S and I vary from 0–1 degree, and H varies from 0° to 360° [23].